US Department of Energy
Office of Environmental Management
National Analytical Management Program

Environmental Management
Electronic Data Deliverable
Tier 1
EMEDD-1
Version 1.3

May 2002

Revised: May 15, 2002

The Internet version of this document is the current official version.

http://emnamp.inel.gov/edd/emedd.html

Environmental Management Electronic Data Deliverable

EMEDD Tier 1

Version 1.3

The Tier 1 Environmental Management Electronic Data Deliverable (EMEDD) was developed to define a common means for analytical laboratories to easily and readily transmit electronic data to a variety of data users. The intent of a Tier 1 deliverable is to provide data of known quality that will assist the data user in making informed decisions. It focuses on data elements that analytical laboratories routinely report electronically, resulting in lower overall project costs and increased efficiency in data evaluation and report generation. The data elements and data table relationships in a Tier 1 deliverable support an unambiguous association between field samples and performance data (method blanks, matrix spikes, etc.), which are directly related to the usability of the data. The Tier 1 deliverable was designed as a base level format that should not be used as the sole basis for determining method or contract compliance. More importantly, the Tier 1 deliverable should not be considered a substitute for good quality assurance.

This document provides the structural and data representation rules for transmitting analytical data following the EMEDD specification. The **Data Element Definitions** section defines the current data elements. The **Entity Relationship Diagram** illustrates the relationships between EMEDD data tables and their corresponding data elements. Within the data tables certain key data elements are required for relational linking of information and must be reported as non-null values. The key data elements for the EMEDD Specification are Lab_Reporting_Batch, Lab_Analysis_ID, Lab_Sample_ID, Batch_ID, and Analyte_ID. Predetermined data value lengths are not defined by this specification. If length restrictions are desired, it is the responsibility of the client to contractually define them.

Two format options have been established for the EMEDD Specification. The **Relational File Format Specification** defines the specific rules for implementing the EDD in a multi-table relational file format. The **Flat File Format Specification** defines the specific rules for implementing the EDD in a single-table flat file format.

Relational File Format

The Relational File Format was designed to transmit analytical data for an EMEDD Tier 1 reporting batch in a single file. Multiple relational tables exist within the file to allow for flexible and efficient reporting.

- The **Data_ Package** table identifies the group of samples reported to the client; the laboratory performing the analysis; the data package transmittal date; and the EDD type and version associated with the data file.

- The **Samples** table identifies the individual samples analyzed that are reported as a group. The table includes the client's field samples and the associated laboratory-generated QC samples. Administrative information related to the samples, such as chain-of-custody data, is reported in this table. The table also includes physical information about the sample as received by the laboratory.

- The **Analyses** table identifies the specific analytical event performed by the laboratory to measure a sample constituent as prescribed by a method.

- The **Results** table identifies the reported analyte; the reportable values of an analysis; limits that may restrict the reportable value; and the laboratory-assigned qualified condition of the analytical value. The Analyte_ID is used to uniquely identify analytes in the Results table.

- The **QC** table identifies method quality control data that is specific to an analyte and used in monitoring a laboratory control process.

- The **Batch** table identifies groups of samples or analyses that are associated by process or reporting requirements. QC_Linkage specifies the batch classification.

  Designated batch values for QC_Linkage:

  - **Analysis Batch:** A laboratory identifier for a batch of analyses done on one instrument associated with the level of detail at which the instrument is checked to be in control.

  - **Clean-up Batch:** A laboratory identifier for a batch of aliquots going through a clean-up step together (i.e. similarity of time, place, and manner of clean-up) as part of preparation for analysis by one method.

  - **Handling Batch:** A laboratory identifier for a batch of samples, not aliquots, handled together (i.e. similarity of time, place, and manner of handling) during the initial processing for analysis by one method.

  - **Preparation Batch:** A laboratory identifier for a batch of aliquots prepared together (i.e. similarity of time, place, and manner of preparation) for analysis by one method. The notion of preparation is used in a generic sense for any activity prior to instrumental analysis. Method blanks and/or laboratory control samples are often used to demonstrate that the lab's process is in control in each preparation batch.

The data contained within the Relational file are of variable length ASCII records. A carriage return with line feed (CR/LF) must separate each record. White space (spaces, tabs, etc.) either preceding or following a data element value within a data element field is ignored. A delimiter shall not follow the last data element of a record.

Each data table consists of at least three data records: one table name record, one data element list record, and one or more detail records.

The table name record identifies the current table by beginning the record with the word "Table", a colon, and the table name (i.e. Table: Data_Package). White space within this record is ignored. The table name record must be the first record in the table. The table name is not case sensitive.

The data element list record contains the data element names in the same order that the data element values occur in the detail records. The data element list record must be the second record, following the table name record. This record specifies the data elements reported in the current table and the sequence of the data element values in the table's detail records. A data element name may not appear more than once in any data element list record. Data element names are not case sensitive and must be separated with an ASCII character 124 (|) as a delimiter. White space within this record is ignored. If the "Comment" data element is used, it is always the last data element in the record.

The data element detail records contain the ordered data element values that correspond to the data element names listed in the data element list record for the current table. Data element values are separated with an ASCII character 124 (|) as a delimiter. Null data values must be reported as a value of zero length with a delimiter to ensure that the sequence of data element values corresponds with the data element list record.

Blank records in the file will be ignored.

Flat File Format

The Flat File Format was designed to transmit analytical data for an EMEDD Tier 1 reporting batch within a single table and single file. The Flat File Format option has been included in this specification to accommodate the most common approach used by the laboratory industry for reporting analytical data. It should be noted, however, that there are some drawbacks associated with the Flat File Format. The disadvantages are related to the fact that several data element values must be repeated throughout the file. The most obvious consequence of repeating information within a file is increased file size. Variation among repeated data element values is another potential concern. Precautionary measures must be taken to ensure that repeated values are consistently reported in the file.

The data contained within the Flat File are of variable length ASCII records. A carriage return with line feed (CR/LF) must separate each record. White space (spaces, tabs, etc.) either preceding or following a data element value within a data element field is ignored. A delimiter shall not follow the last data element of a record.

The first line of the Flat File serves as a data element list record. The data element list record specifies the data element names reported and the order of the data element values in subsequent records. The data elements reported may be a client-specified subset of the EMEDD data elements available. A data element name may not appear more than once in the data element list record. Data element names are not case sensitive, and must be separated with an ASCII character 124 (|) as a delimiter. White space within this record is ignored. If the "Comment" data element is used, it is always the last data element in the record.

Subsequent lines in the Flat File contain the ordered data element values that correspond to the data element list record. Null data values must be reported as a value of zero length with a delimiter to ensure that the sequence of data element values corresponds with the data element list record.

The Batch table data elements specified in the Relational File Format must be reported differently in the Flat File since multiple batch entries will be reported in one record. The following Batch table data elements (Batch_ID, Batch_Type, Batch_Date, Batch_Procedure_ID, Batch_Procedure_Name, and QC_Linkage) must have an underscore and a sequential number appended to the end of the data element to accommodate multiple sample batches (e.g., Batch_ID_1, Batch_Date_1, Batch_Procedure_ID_1, Batch_Type_1, Batch_Procedure_Name_1, and QC_Linkage_1 ... Batch_ID_n ...QC_Linkage_n).

General Format Information

Valid Characters:

Data element values may contain mixed-case standard printing ASCII characters, codes 32 through 126. Restrictions on character sets for data element values may exist because of the specifications of data element data type. The vertical bar, "|", ASCII code 124, is only to be used to separate data elements or their respective data values. Data element values are of variable length, but they may be limited by valid values (current lists posted on the official EMEDD web site), data type assignments or client restrictions.

Date-Time Format:

Data elements defined as date-time values will conform to a two-digit day, three-character month and four-digit year separated with dashes, DD-MMM-YYYY. If the time is required as part of the date value for the data element, the value will contain a two-digit hour, two-digit minute and two-digit second separated with colons, HH:MM:SS. Time is expressed in a 24-hour format with midnight equal to 00:00:00, (ex. 06-May-1999 09:25:30). If values are not available for seconds, then zero fill. Do not report time if it is not available, report date only. Dates are not case sensitive. The Time Zone associated with the date and time of the analytical processes will be the same as the laboratory facility performing the analysis.

Numeric Formats:

Numeric values may be reported in either standard or scientific notation. The value must have 16 or fewer digits and not contain commas. If scientific notation is used, it must be in the format sn.dEsd where (s) is "+" or "-", (n) is a digit and (d) is one or more digits. The general range is 1.0E-39 to 1.0E39. Numeric values must not contain imbedded spaces.

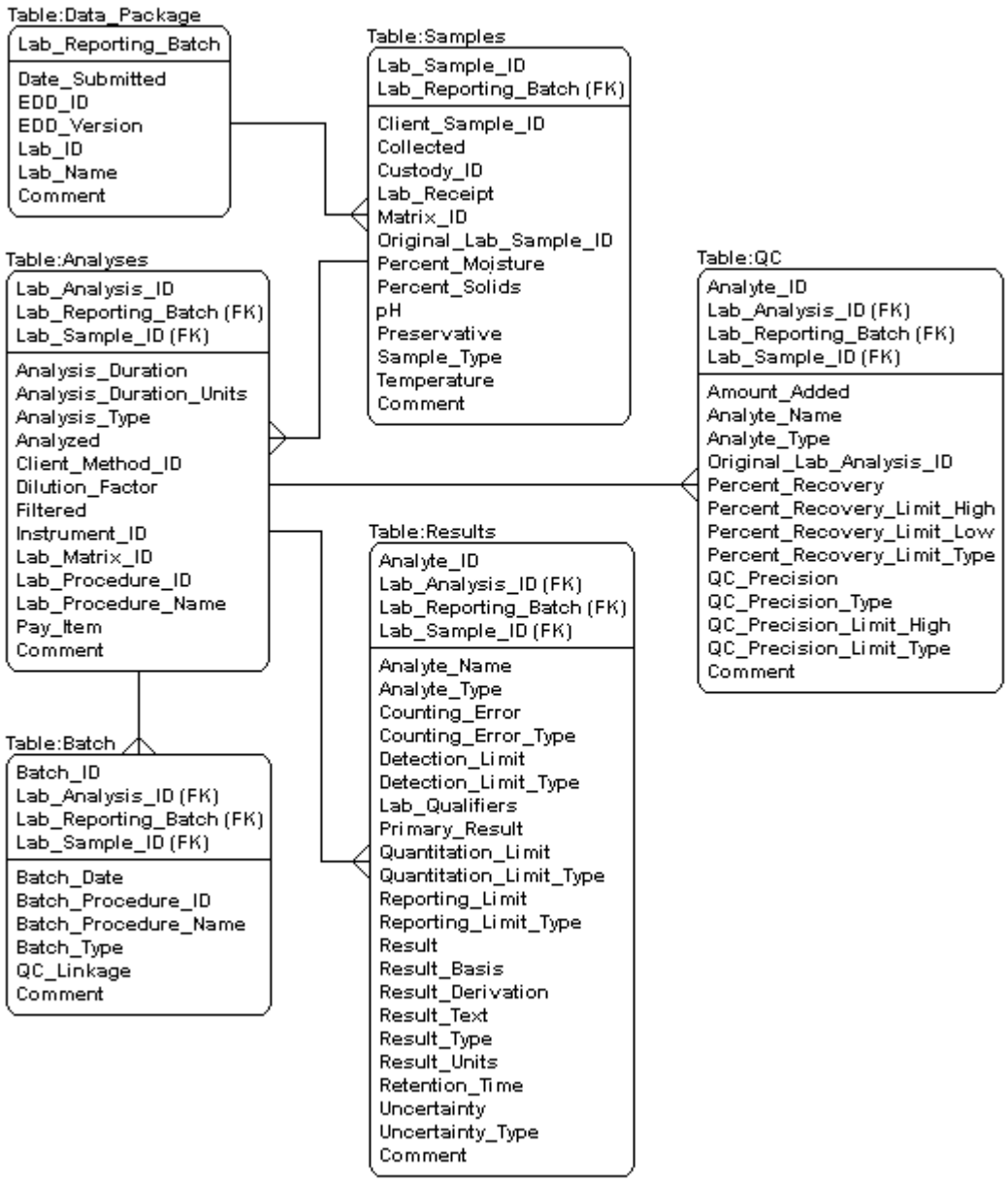
Rounding Rule:

Rounding of values reported in the EDD must be consistent with the values reported on client hardcopy forms. Rounding will be to the number of significant digits specified by the client. If a consistent rounding method is not enforced by the laboratory then the preferred rounding method shall be as follows: If the remainder beyond the last retained digit is greater than 5, increase that digit to the next higher number. If the remainder beyond the last retained digit is less than 5, allow the last digit to remain as it is. If the digit beyond the last retained digit is exactly 5 (with no remainder), round the digit to the nearest even number. If the last retained digit is already even, then it is not changed. If the last retained digit is odd, then add one to make it even. When multiple values are used in calculations, rounding should only be applied to the final value of the calculation.

Units:

Units for the Counting_Error, Detection_Limit, QC_Precision, Quantitation_Limit, Reporting_Limit, and Uncertainty must be expressed in the same units as the Result. Temperature must be reported in degrees Celsius (°C). Retention_Time must be reported in minutes and fraction of minutes.

Entity Relationship Diagram



Data Element Definitions

Amount_Added: Specifies a known amount of analyte that has been added to an aliquot. Is used for spikes, surrogates, and tracers. Expressed as a sample concentration in the same units as the result.
Data Type: Numeric

Analysis_Duration: The duration of the instrumental count time for radiochemical analysis.
Data Type: Numeric

Analysis_Duration_Units: Units for Analysis_Duration.
Valid Values

Analysis_Type: A code used to define the type of analysis**.** This code is also used to identify a single analysis from multiple analyses that are used to generate a single result (e.g., dilutions and re-analyses).
Valid Values

Analyte_ID: Unique identification of a chemical substance, grouping, or sample characteristic. Chemical Abstracts Service (CAS) Registry Numbers are used when that number exists. EPA Chemical Identifiers are used (if available) when CAS numbers do not exist and cannot be assigned. Sample characteristics and chemical substances not published by either CAS or EPA are assigned unique identifiers intended specifically for use with this deliverable. In the case of an unknown TIC, site-specific requirements may be defined. In the absence of site-specific requirements, the Analyte_ID should be reported as UNKNOWN1, UNKNOWN2…for as many TICs as are found in each sample.
Valid Values

Analyte_Name: Laboratory-assigned name for an analyte.
(Maximum field length to be specified by the client)

Analyte_Type: A code that identifies the type of analyte reported.
Valid Values

Analyzed: The date (and time, if required) of analysis of an aliquot.
Data Type: Date-time

Batch_Date: The date (and time, if required) of batch beginning. Required for preparation, clean-up, and handling batches.
Data Type: Date-time

Batch_ID: A laboratory-defined identifier for a group of samples or aliquots that are processed together.
(Maximum field length to be specified by the client)

Batch_Procedure_ID: A laboratory-defined code for the procedure used for sample/aliquot handling, preparation, or cleanup.
(Maximum field length to be specified by the client)

Batch_Procedure_Name: Description of the laboratory's procedure for sample/aliquot handling, preparation, or cleanup.
(Maximum field length to be specified by the client)

Batch_Type: A code used to define the type of batch process used for a group of samples/aliquots**.**
Valid Values

Client_Method_ID: A code for the method used by the laboratory to analyze the sample. This code may be a composite of the method source and the method number. Client-specific codes may be assigned for methods not included in the valid value listing.
Valid Values

Client_Sample_ID: A client-defined identifier for a sample. This should be the basis on which the client identifies the sample. However, not all clients define values for laboratory generated QC samples.
(Maximum field length to be specified by the client)

Collected: The date (and time, if required) the sample was collected. If collected over a range of dates, this is the start date.
Data Type: Date-time

Comment: A free-form comment that can occur in any record.
Data Type: Text

Counting_Error: Error value on counting discrete events, such as is common in radiochemistry.
Data Type: Numeric

Counting_Error_Type: The confidence interval specified by the client for the Counting_Error.
Valid Values

Custody_ID: A client-defined code for sampling documentation, such as a chain-of-custody, associated with receipt of the sample(s) in the laboratory.
(Maximum field length to be specified by the client)

Date_Submitted: The date the data package was delivered to the client.
Data Type: Date-Time

Detection_Limit: Detection limit for the analyte being measured. Detection limits are defined in terms of the presence or absence of the analyte.
Data Type: Numeric

Detection_Limit_Type: Specifies the type of detection limit. The client and the laboratory should agree upon the precise definition of the Detection_Limit_Type reported.
Valid Values

Dilution_Factor: The overall dilution of the aliquot. A value of one corresponds to nominal conditions for the method. Values greater than one correspond to dilutions. Values less than one correspond to concentrations.
Data Type: Numeric

EDD_ID: Identifies the format specification used to create the electronic data deliverable.
Valid Values

EDD_Version: Identifies the version of the format specification used to create the electronic data deliverable.
Valid Values

Filtered: Indicates whether or not a sample/aliquot was filtered in the laboratory. Reported as either "YES" or "NO".

Instrument_ID: A laboratory-defined code for an instrument. Should be unique within the laboratory.
(Maximum field length to be specified by the client)

Lab_Analysis_ID: Unique laboratory identifier for a single run of a single aliquot, single instrument, and single detector. The Lab_Analysis_ID will typically identify an analysis at a moment in time (e.g., a metals analysis encompassing different preparations and technologies for one sample will require a unique Lab_Analysis_ID for each instrument injection). The Lab_Analysis_ID may also be used to help relate events not addressed in a Tier 1 deliverable that contribute to a reported result (e.g., when a result is computed from multiple measurements).
(Maximum field length to be specified by the client)

Lab_ID: Identifier for the laboratory performing the analysis. Client-specific codes may be assigned for on-site laboratories.
Valid Values

Lab_Matrix_ID: Identifier for the matrix of an aliquot as reported by the laboratory. Should be compatible with what is reported for Matrix_ID unless a handling or preparation process alters the original sample matrix.
Valid Values

Lab_Name: Descriptive name for the laboratory performing the analysis.
(Maximum field length to be specified by the client)

Lab_Procedure_ID: A laboratory-defined code for the procedure used for the analytical method.
(Maximum field length to be specified by the client)

Lab_Procedure_Name: Description of the laboratory procedure used for the analytical method.
(Maximum field length to be specified by the client)

Lab_Qualifiers: A string of single letter result qualifiers assigned by the laboratory, based on defined rules and values.
Valid Values

Lab_Receipt: The date (and time, if required) the sample was received in the laboratory.
Data Type: Date-time

Lab_Reporting_Batch: A laboratory-defined identifier for a batch of samples reported as a group by the laboratory. Rules for determining the Lab_Reporting_Batch may be specified by the client. The Lab_Reporting_Batch is also commonly referred to as the Sample Delivery Group (SDG) or Report Data Group (RDG).
(Maximum field length to be specified by the client)

Lab_Sample_ID: Laboratory identifier for a sample that uniquely identifies a single sample that is subjected to an analysis. This code is the primary link into the laboratory's record keeping system. It is not necessarily one-to-one with the Client_Sample_ID. Should be unique within the laboratory.
(Maximum field length to be specified by the client)

Matrix_ID: A code for the sample matrix as received by the laboratory, prior to any handling or preparation procedures. Should be compatible with what is reported on a chain-of-custody form.
Valid Values

Original_Lab_Analysis_ID: The Lab_Analysis_ID of the original analysis from which the QC sample result was derived.
(Maximum field length to be specified by the client)

Original_Lab_Sample_ID: The Lab_Sample_ID of the original regular sample from which the QC sample was derived.
(Maximum field length to be specified by the client)

Pay_Item: Client-specified identifier defining an analysis request.
(Maximum field length to be specified by the client)

Percent_Moisture: Percent of sample composed of water.
Data Type: Numeric

Percent_Recovery: The recovery of an analyte expressed as a percentage of the amount added.
Data Type: Numeric

Percent_Recovery_Limit_High: The upper limit for the Percent_Recovery. Units are the same as for Percent_Recovery.
Data Type: Numeric

Percent_Recovery_Limit_Low: The lower limit for the Percent_Recovery. Units are the same as for Percent_Recovery.
Data Type: Numeric

Percent_Recovery_Limit_Type: The source for the Percent_Recovery limits.
Valid Values

Percent_Solids: Percent of the sample composed of solid material.
Data Type: Numeric

---

pH: The negative of the logarithm of the hydrogen ion potential.
(Maximum field length to be specified by the client)

Preservative: Preservative added to the sample.
(Maximum field length to be specified by the client)

Primary Result: Indicates whether or not a result is selected by the laboratory to be the reportable result (e.g., Primary_Result identifies the reportable result when multiple dilutions or re-analyses are performed). Reported as either "YES" or "NO".

QC_Linkage: For a method QC sample, specifies which batch is the basis for the association between the QC sample and the regular samples.
Valid Values

QC_Precision: A calculated quality control value used as an indicator of precision for the results of an analytical procedure.
Data Type: Numeric

QC_Precision_Limit_High: The upper limit for the reported QC_Precision. Units are the same as for QC_Precision.
Data Type: Numeric

QC_Precision_Limit_Type: The source for the reported QC_Precision limit.
Valid Values

QC_Precision_Type: The intended use for the reported QC_Precision.
Valid Values

Quantitation_Limit: Quantitation limit for the analyte being measured. Quantitation limits are defined in terms of a specified degree of uncertainty for results at this level.
Data Type: Numeric

Quantitation_Limit_Type: One of a list of client-defined acronyms that specify the type of quantitation limit.
Valid Values

Reporting_Limit: Reporting limit for the analyte being measured. Reporting limits are defined in terms of a value below which data is reported as not detected.
Data Type: Numeric

Reporting_Limit_Type: One of a list of client-defined acronyms that specify the type of reporting limit.
Valid Values

Result: Reportable result in numeric format for an analyte. The result may be computed from multiple measurements.
Data Type: Numeric

Result_Basis: The basis upon which the results were calculated.
Valid Values

Result_Derivation: Indicates how the reported result was derived (e.g., single measurement, summation of values, etc.). Note: The laboratory and the client should have an understanding of the laboratory's convention for deriving results when the derivation process is not prescribed in the analytical method.
Valid Values

Result_Text: Reportable result containing non-numeric characters for an analyte (e.g., the reported result for coliform bacteria may be reported as "ABSENT" or "PRESENT").
(Maximum field length to be specified by the client)

Result_Type: Specifies whether the reported result is qualitative, quantitative, semi-quantitative, or non-numeric.
Valid Values

Result_Units: Units for Results.
Valid Values

Retention_Time: The time between injection and detection for a target analyte using chromatography or other techniques.
Data Type: Numeric

Sample_Type: Client-defined code for the type of sample.
Valid Values

Temperature: The temperature of the sample as received.
Data Type: Numeric

Uncertainty: Total propagated uncertainty.
Data Type: Numeric

Uncertainty_Type: Specification for the uncertainty.
Valid Values
